DOCUMENT RESUME

ED 051 281 TM 000 597

AUTHOR	Clark, Phillip I.; And Others
TITLE	The Use, Misuse, and Abuse of Tests.
INSTITUTION	New England Association for Measurement and Evaluation in Guidance.
SPONS AGENCY	Harcourt Brace Jovanovich, Inc., New York, N.Y.; Rhode Island State Agency for Elementary and Secondary Education, Providence.
PUB DATE	May 70
NOTE	35p.; Report on the First Annual NEAMEG Conference on Measurement in Education, Newport, Rhode Island, May 14-15, 1970
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Computer Oriented Programs, *Conference Reports, Confidentiality, Criterion Referenced Tests, *Educational Accountability, Educationally Disadvantaged, Federal Programs, *Measurement, National Programs, Norms, Relevance (Education), State Surveys, *Testing, *Test Interpretation, Tests
IDENTIFIERS	Jensen Report, *National Assessment

ABSTRACT

The New England Association for Measurement and Evaluation in Guidance (NEAMEG) Conference on Measurement in Education was designed to (1) provide a forum for the examination and discussion of vital issues related to measurement and evaluation; (2) facilitate communication among educators from various disciplines and levels of education within the New England region, and to encourage their active involvement in "attacking" current identified problems and concerns relating to the use of tests and other evaluative devices; and (3) stimulate the development of a series of position papers stating the views of the professional members of the NEAMEG as a group, which may serve as guidelines for education. The proceedings include: "Innovative Test Usage for Individual Pupil Growth," Philip I. Clark; "National Assessment," Thomas R. Knapp; "State Testing Programs," Paul B. Campbell; "Testing the Disadvantaged," Lenore A. DeLucia; "Computerization in Relation to Testing and Evaluation," James R. Baker; "Testing and its Relevancy to the Seventies," Thomas Burns; "Federally Funded Programs," Thomas Burns; "Disclosure of Test Results," Thomas P. Nally; "Norms: Fact or Fancy," Walter N. Durost; "Tests: Who or What is Being Evaluated," C. Thomas Skoggs; and "The Jensen Report," Paul B. Campbell. A summary of the discussion by the reactors to each presentation follows each paper. (DG)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Report on:
The First Annual NEAMEG Conference
on Measurement in Education
Newport, R. I. / May 14-15, 1970

The Use, Misuse, and Abuse of Tests

ED051281

Sponsors: New England Association for Measurement and Evaluation in Guidance; Rhode Island State Agency for Elementary and Secondary Education (Titles I, ESEA, and V-A, NDEA), and Test Department, Harcourt, Brace & World, Inc.

TM 000 592

1a ::::: b ::::: c ::::: d ::::: e :::::
2a ::::: b ::::: c ::::: d ::::: e :::::
3a ::::: b ::::: c ::::: d ::::: e :::::
4a ::::: b ::::: c ::::: d ::::: e :::::
5a ::::: b ::::: c ::::: d ::::: e :::::
6a ::::: b ::::: c ::::: d ::::: e :::::
7a ::::: b ::::: c ::::: d ::::: e :::::
8a ::::: b ::::: c ::::: d ::::: e :::::
9a ::::: b ::::: c ::::: d ::::: e :::::
10a ::::: b ::::: c ::::: d ::::: e :::::
11a ::::: b ::::: c ::::: d ::::: e :::::
12a ::::: b ::::: c ::::: d ::::: e :::::
13a ::::: b ::::: c ::::: d ::::: e :::::
14a ::::: b ::::: c ::::: d ::::: e :::::
15a ::::: b ::::: c ::::: d ::::: e :::::
16a ::::: b ::::: c ::::: d ::::: e :::::
17a ::::: b ::::: c ::::: d ::::: e :::::
18a ::::: b ::::: c ::::: d ::::: e :::::
19a ::::: b ::::: c ::::: d ::::: e :::::
20a ::::: b ::::: c ::::: d ::::: e :::::
21a ::::: b ::::: c ::::: d ::::: e :::::
22a ::::: b ::::: c ::::: d ::::: e :::::
23a ::::: b ::::: c ::::: d ::::: e :::::
24a ::::: b ::::: c ::::: d ::::: e :::::
25a ::::: b ::::: c ::::: d ::::: e :::::
26a ::::: b ::::: c ::::: d ::::: e :::::
27a ::::: b ::::: c ::::: d ::::: e :::::
28a ::::: b ::::: c ::::: d ::::: e :::::
29a ::::: b ::::: c ::::: d ::::: e :::::
30a ::::: b ::::: c ::::: d ::::: e :::::
31a ::::: b ::::: c ::::: d ::::: e :::::
32a ::::: b ::::: c ::::: d ::::: e :::::
33a ::::: b ::::: c ::::: d ::::: e :::::
34a ::::: b ::::: c ::::: d ::::: e :::::
35a ::::: b ::::: c ::::: d ::::: e :::::
36a ::::: b ::::: c ::::: d ::::: e :::::
37a ::::: b ::::: c ::::: d ::::: e :::::
38a ::::: b ::::: c ::::: d ::::: e :::::
39a ::::: b ::::: c ::::: d ::::: e :::::
40a ::::: b ::::: c ::::: d ::::: e :::::
41a ::::: b ::::: c ::::: d ::::: e :::::
42a ::::: b ::::: c ::::: d ::::: e :::::
43a ::::: b ::::: c ::::: d ::::: e :::::
44a ::::: b ::::: c ::::: d ::::: e :::::
45a ::::: b ::::: c ::::: d ::::: e :::::
46a ::::: b ::::: c ::::: d ::::: e :::::
47a ::::: b ::::: c ::::: d ::::: e :::::
48a ::::: b ::::: c ::::: d ::::: e :::::
49a ::::: b ::::: c ::::: d ::::: e :::::
50a ::::: b ::::: c ::::: d ::::: e :::::
51a ::::: b ::::: c ::::: d ::::: e :::::
52a ::::: b ::::: c ::::: d ::::: e :::::
53a ::::: b ::::: c ::::: d ::::: e :::::
54a ::::: b ::::: c ::::: d ::::: e :::::
55a ::::: b ::::: c ::::: d ::::: e :::::
56a ::::: b ::::: c ::::: d ::::: e :::::
57a ::::: b ::::: c ::::: d ::::: e :::::
58a ::::: b ::::: c ::::: d ::::: e :::::
59a ::::: b ::::: c ::::: d ::::: e :::::
60a ::::: b ::::: c ::::: d ::::: e :::::
61a ::::: b ::::: c ::::: d ::::: e :::::
62a ::::: b ::::: c ::::: d ::::: e :::::
63a ::::: b ::::: c ::::: d ::::: e :::::
64a ::::: b ::::: c ::::: d ::::: e :::::
65a ::::: b ::::: c ::::: d ::::: e :::::
66a ::::: b ::::: c ::::: d ::::: e :::::
67a ::::: b ::::: c ::::: d ::::: e :::::
68a ::::: b ::::: c ::::: d ::::: e :::::
69a ::::: b ::::: c ::::: d ::::: e :::::
70a ::::: b ::::: c ::::: d ::::: e :::::
71a ::::: b ::::: c ::::: d ::::: e :::::
72a ::::: b ::::: c ::::: d ::::: e :::::
73a ::::: b ::::: c ::::: d ::::: e :::::
74a ::::: b ::::: c ::::: d ::::: e :::::
75a ::::: b ::::: c ::::: d ::::: e :::::
76a ::::: b ::::: c ::::: d ::::: e :::::
77a ::::: b ::::: c ::::: d ::::: e :::::
78a ::::: b ::::: c ::::: d ::::: e :::::
79a ::::: b ::::: c ::::: d ::::: e :::::
80a ::::: b ::::: c ::::: d ::::: e :::::
81a ::::: b ::::: c ::::: d ::::: e :::::
82a ::::: b ::::: c ::::: d ::::: e :::::
83a ::::: b ::::: c ::::: d ::::: e :::::
84a ::::: b ::::: c ::::: d ::::: e :::::
85a ::::: b ::::: c ::::: d ::::: e :::::
86a ::::: b ::::: c ::::: d ::::: e :::::
87a ::::: b ::::: c ::::: d ::::: e :::::
88a ::::: b ::::: c ::::: d ::::: e :::::
89a ::::: b ::::: c ::::: d ::::: e :::::
90a ::::: b ::::: c ::::: d ::::: e :::::
91a ::::: b ::::: c ::::: d ::::: e :::::
92a ::::: b ::::: c ::::: d ::::: e :::::
93a ::::: b ::::: c ::::: d ::::: e :::::
94a ::::: b ::::: c ::::: d ::::: e :::::
95a ::::: b ::::: c ::::: d ::::: e :::::
96a ::::: b ::::: c ::::: d ::::: e :::::
97a ::::: b ::::: c ::::: d ::::: e :::::
98a ::::: b ::::: c ::::: d ::::: e :::::
99a ::::: b ::::: c ::::: d ::::: e :::::
100a ::::: b ::::: c ::::: d ::::: e :::::
101a ::::: b ::::: c ::::: d ::::: e :::::
102a ::::: b ::::: c ::::: d ::::: e :::::
103a ::::: b ::::: c ::::: d ::::: e :::::
104a ::::: b ::::: c ::::: d ::::: e :::::
105a ::::: b ::::: c ::::: d ::::: e :::::
106a ::::: b ::::: c ::::: d ::::: e :::::
107a ::::: b ::::: c ::::: d ::::: e :::::
108a ::::: b ::::: c ::::: d ::::: e :::::
109a ::::: b ::::: c ::::: d ::::: e :::::
110a ::::: b ::::: c ::::: d ::::: e :::::
111a ::::: b ::::: c ::::: d ::::: e :::::
112a ::::: b ::::: c ::::: d ::::: e :::::
113a ::::: b ::::: c ::::: d ::::: e :::::
114a ::::: b ::::: c ::::: d ::::: e :::::
115a ::::: b ::::: c ::::: d ::::: e :::::
116a ::::: b ::::: c ::::: d ::::: e :::::
117a ::::: b ::::: c ::::: d ::::: e :::::
118a ::::: b ::::: c ::::: d ::::: e :::::
119a ::::: b ::::: c ::::: d ::::: e :::::
120a ::::: b ::::: c ::::: d ::::: e :::::
121a ::::: b ::::: c ::::: d ::::: e :::::
122a ::::: b ::::: c ::::: d ::::: e :::::
123a ::::: b ::::: c ::::: d ::::: e :::::
124a ::::: b ::::: c ::::: d ::::: e :::::
125a ::::: b ::::: c ::::: d ::::: e :::::
126a ::::: b ::::: c ::::: d ::::: e :::::
127a ::::: b ::::: c ::::: d ::::: e :::::
128a ::::: b ::::: c ::::: d ::::: e :::::

PLANNING COMMITTEE

CHAIRMAN:
Miss Kathryn R. Orme
Supervisor of Psychological Services,
School Department, Providence, R.I.

CO-CHAIRMAN:
Miss Elizabeth A. Casey
Guidance Specialist
School Department, Providence, R.I.

THE COMMITTEE:
Mrs. Eleanor F. Cates
Test Department, Harcourt, Brace & World, Inc.

Mrs. Rita F. Clark
Director of Guidance, Tolman High School, Pawtucket, R.I.

Mr. John DeLuca
Counselor, Portsmouth, R.I.,School Department

Dr. Frances E. Dunn
Director of Educational Measurement
Brown University, Providence, R.I.

Miss Eileen A. Matteo
State Consultant, Elementary Guidance
Rhode Island State Agency for Elementary and Secondary Education

Dr. Thomas P. Nally
Department of Education
University of Rhode Island, Kingston, R.I.

Miss Barbara Sutcliffe
Psychological Examiner, Pawtucket, R.I., School Department

Mr. Charles H. Tobin
Director of Guidance
Rogers Hill School, Newport, R.I.

Mr. Robert W. Trager
Counselor, Bristol, R.I., Junior-Senior High School

Dr. Joseph A. Whelan
Curriculum Director, North Providence, R.I., School Department

Mrs. E. Natalie Williams
Director of Guidance, Barrington, R.I., Junior High School

ED051281

TABLE OF CONTENTS

FOREWORD

3 Measurement - Some Implications for the 70's?
By Miss Eileen A. Matteo

PRESENTATIONS AND PANELS

Opening Address:

6 Innovative Test Usage for Individual Growth
By Dr. Phillip I. Clark

Panels and Discussions - Testing: Applications and Practices:

8 1. National Assessment
Dr. Thomas R. Knapp, presenter

10 2. State Testing Programs
Dr. Paul B. Campbell, presenter

12 3. Testing the Disadvantaged
Dr. Lenore A. DeLucia, presenter

14 4. Computerization
Dr. James R. Baker, presenter

Dinner Address:

17 Testing and its relevancy to the Seventies
By Thomas Burns

Panels and Discussions - Testing: Issues and Policies:

19 1. Federally Funded Programs
Thomas Burns, presenter

22 2. Disclosure of Test Results
Dr. Thomas P. Nally, presenter

25 3. Norms: Facts or Fancy
Dr. Walter N. Durost, presenter

27 4. Tests: Who or What is Being Evaluated
Dr. C. Thomas Skaggs, presenter

Special Issue Panel:

30 The Jensen Report
Dr. Paul B. Campbell, presenter

FOREWORD:
MEASUREMENT – SOME IMPLICATIONS FOR THE 70's?

By EILEEN A. MATTEO
Consultant, Elementary Guidance
Rhode Island State Agency for Elementary and Secondary Education

On May 14 and 15, 1970 representatives of the state education agencies, various institutions of higher education and local education agencies of the six New England States attended an Invitational Conference on Measurement in Education which had as its theme, "The Use, Misuse, and Abuse of Tests". This conference, for which Rhode Island served as Host State, represented the first of the annual spring conferences to be convened by the New England Association for Measurement and Evaluation in Guidance and emphasized practical application of the issues explored. The New England Association for Measurement and Evaluation in Guidance is a newly organized affiliate of the national Association for Measurement and Evaluation in Guidance which is a division of the American Personnel and Guidance Association. Sponsorship was shared with the Test Department of Harcourt, Brace and World, Inc. and the Rhode Island State Agency for Elementary and Secondary Education (Titles I, ESEA and V-A, NDEA).

The conference was designed to: (1) provide a forum for the examination and discussion of some of the most vital issues related to measurement and evaluation as they effect education today; (2) facilitate communication among and bring together for more active involvement educators from a variety of disciplines and from all levels of education within the New England region as a means of "attacking" current identified problems and concerns relating to the use of tests and other evaluative devices in the field of education; and (3) stimulate the development of a series of position papers stating the point of view of the professional members of the New England Association for Measurement and Evaluation in Guidance as a group, which may serve as guidelines for the field of education.

The conference program touched upon several "burning" issues in the area of measurement and evaluation as evidenced by discussion of such questions as: What effects or results will national assessment have as an inventory of knowledge, skill, and attitude? What to test? Whom to test? When to test? How much to test? How to utilize the test information? What are the opportunities for testing to improve education? What need is there to improve the use of test results? What need exists for devising new means of reporting test results? How can the potential and achievement of disadvantaged youth be best evaluated? Can test results be reported in a better way? What are the advantages and disadvantages of various kinds of norms? What is the meaning of testing for accountability in education? To whom and how should test results be disclosed?

Recurring Issues and Themes

Emerging from the conference activities were several recurring issues and themes. The following reflect some of those which may require particular consideration and investigation by professional educators:

1. Such projects as National Assessment which typify renewed concern or questions about the effectiveness of our schools and their programs in meeting the real educational needs of all youth.

2. The type of test content that can be validated in diagnosing the educational needs of disadvantaged children.

3. The increasingly significant role of the computer in more effectively acquiring and using data about pupils, as well as in changing instructional processes and personnel.

4. The better use of test results for evaluative purposes. While tests have been used in the past to evaluate students, there is evidence of the need for such instruments to be used to evaluate teaching practices, curriculum, and organizational factors. Therefore, how test results might be used for such evaluative purposes is in need of further definition.

5. As measurement will always be with us, greater need for education of the public exists in terms of understanding the real meanings of tests instead of catering to common biases.

6. As accountability in education is at the forefront, the role and functions of school boards, as policy makers, take on heightened importance, e.g., their strong investigative and appraising function.

7. The eternal question of environment versus heredity perpetuates and thus, raises the dual question of what is the nature of the child when he comes to us and how can we best work with this as a "given".

Recommendations for Action

As a result of the formal presentations and lively small group discussions, the conference also produced, in part, the following implications for immediate study and action:

1. That the New England Association for Measurement and Evaluation in Guidance:

   a. Develop and publish a policy statement concerning the disclosure and use of test results which can serve as a position to be adopted by state and local education agencies within the New England region.

   b. Establish and provide guidelines to aid educational personnel within the New England area as to the proper use and interpretation of objective data in guiding and aiding the educational development of each individual pupil.

2. That New England professional personnel at the local education level use test results and other evaluative methods for improvement of curriculum and instructional practices.

3. That pupil personnel specialists conceptualize their role as including curriculum development and revision.

In summary, the emphasis in the conference program was on the providing of practical information and suggestions from various points of view and levels of authority. Among the program speakers and reactors were included nationally recognized test and measurement experts, college professors and researchers in psychology and education, school administrators, and representatives of the Federal educational agency and several State educational agencies. The proceedings which appear on the following pages provide summaries of the keynote and general session addresses as well as of the speeches and reports presented at the various small discussion groups of the conference. In addition, the reactions, discussions, and/or implications for action that these presentations engendered at the conference are mirrored at the end of each summary.

It is the hope of the New England Association for Measurement and Evaluation in Guidance that this publication will be of value to you in your professional work.

Sincere gratitude is expressed to Dr. Robert W. Read, Publications Committee Chairman, for his diligent work that makes the publication of these proceedings possible.

INNOVATIVE TEST USAGE FOR INDIVIDUAL PUPIL GROWTH

By DR. PHILIP I. CLARK
Chief Consultant to Education, Programs and Services,
Test Department, Harcourt, Brace & World, Inc.

"What is innovation?" An "innovation" does not have to be new. If educators would confront the unequivalence between their expectations and student performance, they would become truly innovative. Innovation in test usage requires a solid working knowledge of testing and evaluation.

Testing instruments are altered as education changes. Test publishers follow changes in instructional materials, curriculum, methods, and philosophy. While publishers may have ideas for new or altered materials, they ordinarily will not be implemented until there is a public to demand them. For example, some test authors are interested in eliminating the grade equivalent score but the test users seem to require such a measure. The shift in many schools to a non-graded system is one example of an educational change that must affect the use of tests. "If grade is to be diminished in importance as a basis for norms within the next decade or eliminated altogether, then obviously we must substitute something else as a basis for grouping students."

One possibility is the mastery test. Mastery is a method of organizing discrete educational objectives which are meaningful to the individual. Items in such tests must be quite different from traditional items, for they measure stated objectives of instruction, specific for time of a school course or lesson. In a mastery test, the score is absolute - a predetermined number of correct responses is required to indicate command of the subject matter. Each item is more important than the mean score on any particular topic.

Another crucial question is the measurement of growth. Can we find new ways to measure growth? For example, in certain special areas, how do we measure growth in a child's self-concept in a Title I program?

There are two other important questions. How are we going to deal with increased pressures for teacher accountability? Will computer-assisted instruction change the role of the teacher from instructor to social group leader and discussion monitor?

Some current testing trends are: (1) Portland, Oregon has a test item bank, allowing for retrieval of items by subject matter and grade level; (2) There is a growing feeling that readiness tests are better than tests of mental ability at applicable, lower grade levels; (3)There is a strong movement away from IQ per se; (4) True item analyses are more in demand; (5) Testing in lower and higher grades is becoming more and more acceptable; (6) There is a new emphasis on listening in reading; (7) The Instructional Placement Report of the Stanford Diagnostic Reading Test gives grouping of children according to similar profiles on subtests; (8) Foreign language achievement is being tested at levels rather than grades, (9) The Ohio Vocational Interest Survey has been designed to relate results of a questionnaire and scores on 24 interests defined in terms of the world of work.

Questions from the floor prompted four recommendations: (1) Testing children in non-graded situations requires special selection and interpretation of tests. (2) Educators must be more concerned with rate of growth than with achievement at any given point in time. (3) Test users must become more sophisticated in their knowledge and application of materials. (4) Test users must begin, as consumers, to demand of the publishers the kinds of instruments they need.

NATIONAL ASSESSMENT

By DR. THOMAS R. KNAPP
Associate Professor, College of Education, University of Rochester

The notion of National Assessment arose because "We do not have product measures in education." The goal of National Assessment is not to compare one school to another, or one system to another, but to measure the use and success of the nation's educational resources "in the broadest sense." It is similar to a census, but differs in two important ways: (1) every exercise has a right answer and (2) a sample of the population is used.

To assess the country's educational progress, the nation has been divided into four geographical areas (Northeast, South, North Central and West) and the population also has been divided into four categories (Large Central City, Urban Fringe, Middle-Sized City, and Rural or Small City). Two populations (school districts and census tracts) have been sampled in order to assess nine, thirteen, and seventeen year olds and also young adults (twenty-six to thirty-five). The samples have been further stratified by sex, color and socio-educational status. Out of ten subject matter areas, two to four will be sampled each year. These ten areas are: Art, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies, Career and Occupational Development and Writing. In July of 1970, results will be available for the first assessment. These first results will give a baseline for the assessment of Citizenship, Science, and Writing. Areas will be reassessed in a cyclical pattern. Some areas (e.g. Science and Reading) may require more frequent assessment than others. In order to be included, objectives and exercises in each area had to be considered important by subject matter experts, within the province of the schools, and relevant to the layman.

Opposition to National Assessment has come from those fearing that it will result in loss of local control of education, but opponents do not fully understand its objective: to take a broad look at the educational product of the nation and to determine progress, stability, or regress in knowledge, skills and attitudes in ten subject matter areas.

Discussion

Moderator: Ruth L. Pennell
State Guidance Consultant
Maine State Department of Education

Reactors: Russell A. Burnham
Executive Secretary
Rhode Island Association of School Committees

Catherine Corcoran
Superintendent of Schools
Southwestern Vermont Union School District

Harry S. Westcott
Superintendent of Schools
Scituate, Rhode Island

Both Mr. Burnham and Mr. Westcott agreed that education must become more business-like. It must begin to report its progress to the public. National Assessment "provides a vehicle for a comprehensive report to the stock holders in the next decade." In order to justify federal, state, and local budgetary increments, educators must be able to show that they are doing what they mean to do. And, they must be able to define in laymen's terms what it is that they mean to do. Mr. Burnham emphasized that a school board, a policy-making body, should have a strong investigative function. It must be willing to appraise its programs.

The reactors worried about the possible, and perhaps inevitable, mis-interpretation of the results of the National Assessment. As Dr. Knapp mentioned, comparison between states is not justified because there are no state boundaries in the National Assessment design. But the press and school systems along with parent groups may be eager to draw such comparisons. "Our job is to gather and report raw data," said Dr. Knapp. The results will be given in the most general terms, i.e., "Ninety per cent of the 17 year olds from small rural settings in the Northeast can write a paragraph at least as good as . . . . "

The reactors and the audience were concerned about the "young adult" sample. Who is tested? Those at home and perhaps unemployed? What part does willingness to participate play in the findings?

The areas that are evaluated have to be flexible to allow for changes in American life. For example, we may have to re-evaluate what mathematics skills are necessary in the age of computers.

Dr. Knapp closed by recommending highly a book written by Frank Womer, Staff Director of the National Assessment Project, and published by the National Assessment office in Ann Arbor, Michigan as an excellent description of the objectives and design of National Assessment.

STATE TESTING PROGRAMS

By DR. PAUL B. CAMPBELL
Office of Research and Statistics,
Pennsylvania State Department of Public Instruction

The functions of state testing programs are usually a compromise between two goals: (1) assessment of institutional effectiveness, and (2) diagnosis, assessment, and prediction of individual performance. Standardized tests are the most commonly used tools for both purposes. States may provide a scoring service or an advisory service. The "in" idea is a provision for a comprehensive assessment program. The institutional type of assessment seems to be emerging as the main state concern because individual diagnosis and prediction can better be handled at the local level.

The Pennsylvania Plan for the Assessment of Educational Quality illustrates one state testing program. The State Legislature of Pennsylvania passed an act requiring the development of assessment techniques and allocated funds for this work. In 1965, a committee developed ten goals of quality education: (1) Development of the self-concept; (2) Understanding and appreciation of ethnic, religious and racial differences; (3) Mastery of use of words and numbers; (4) Good citizenship; (5) Enjoyment of the learning process; (6) Good health habits; (7) Opportunity to be creative; (8) Vocational development; (9) Appreciation of human achievement in natural and social sciences, the arts, and the humanities; (10) Preparation for a world of rapid change.

The purpose of Phase I was to determine the extent to which students were fulfilling these goals. Four objectives were set: (1) Measurement instruments were developed, tested, and refined; (2) Data collection procedures for gathering information on school and community conditions were developed, tested and refined; (3) Hypotheses concerning the relation of community conditions and pupil achievement were tested and analyzed; and (4) Computer analytic techniques were developed, tried out, and refined. The findings of Phase I indicated "those factors which pupils bring with them - levels of previous learning and educational and occupational levels of parents - are most significant in determining how well pupils achieve." More startling is the finding that in many of the goals, less than half of the differences in pupil achievement is accounted for by these factors; schools can and do make a difference.

Phase II was designed to: (1) Provide patterns of student performance on each of the ten goals; (2) Reconfirm the hypotheses; (3) Provide regression weights for use in subsequent phases; and (4) Establish adequacy of the measuring instruments.

Under Phase III which begins in the Fall of 1970, school personnel will be able to implement the developments of Phases I and II. "Assessment will be in terms of student performance, with differences among individuals, schools, and communities taken into account." Four analyses and reporting

procedures will characterize Phase III: (1) Prediction of Expected School Means; (2) Comparison of Expected and Actual Means; (3) Comparison of Expected Student Distributions with Actual Student Distributions; and (4) Evaluation of Student Responses to Key Items. Where serious discrepancies between expected and actual performance exist, school personnel will be able to modify programs. Where predicted and actual achievement agree, parents and teachers may wish to move out of their expected output category. Where discrepancies are positive, local and state personnel will ascertain the contributing factors so that other schools may benefit.

Discussion

Moderator: Dr. Grace M. Glynn
Associate Director
Rhode Island State Agency
for Elementary and Secondary Education

Reactors: Paul E. Campbell
Consultant, Testing Services
Rhode Island State Agency
for Elementary and Secondary Education

Dr. Donald R. Randall
Director of Research and Testing
New Hampshire State Department of Education

In the discussion it was brought out that State testing programs were encouraged by the availability of data banks, computers able to handle eighteen or nineteen variables simultaneously, legislative pressure, and the employment of research and development people by the State Departments and/or Agencies of Education. In New Hampshire, data are being accumulated on Independent Variables (socio-economic status, school budgets, school staffs, etc.) and Dependent Variables (student performance). The most influential of the Independent Variables are in the realm of what school committees have the power to change.

Reactors were interested in discovering how legislatures become enlightened enough to pass such laws! It seems clear that although most of the funding now is federal, states will have to begin to support such programs.

Problems of administration were raised. The reactors and audience members expressed concern regarding teacher misadministration of tests, slow return of test results, and the general hostility to state testing programs. In-service training of testers on the local level and personal contact between the local coordinator (appointment by the superintendent), the state department representative, and the school personnel contributed to the success of the Pennsylvania Plan. The Pennsylvania Plan did not result in the ranking of schools.

TESTING THE DISADVANTAGED

By DR. LENORE A. DELUCIA
Professor of Psychology, Rhode Island College

The goal of an evaluation is to determine if a particular educational program is successful. But how do we define success? Formerly, evaluation of programs was voluntary; with the advent of federal legislation designed to aid the disadvantaged, evaluation was mandated. Evaluation has, for the most part, been program-oriented. Has the reading level of a group been raised by participation in a certain program? In Rhode Island, the State Agency has moved from a program-orientation to a child-orientation in its evaluative procedures. "Compensatory programs are only successful if the children in them are successful." The State Agency worked to help those closest to the child, the teachers, to focus their attention on the individual child.

Unfortunately, the evaluative needs of the federal and state governments are not often congruent with the needs of local school systems, the classrooms, and individual children. Traditional achievement tests were administered for evaluating success in certain objectives of the compensatory programs, i.e. reading. But "We know there are other ways to measure behavior and other behaviors to measure."

To measure the successful achievement of other objectives, the unobtrusive instruments must be developed at the local level. Only the local personnel are able to determine success criteria for their particular children, but "There seems to be a feeling of incompetence among teachers to develop locally made evaluation techniques. The farther the origin of the evaluation instrument from the local level, the more validity it is granted. Nothing could be farther from the truth." A teacher is limited in designing unobtrusive instruments to measure success only by the limits of his or her imagination. Does the child take out more books from the library? Is he tardy less often? Absent less often?

Local evaluation must serve a feedback function. Information must go back to program administrators and designers. "The only justification for evaluation programs is to improve them." Therefore, Dr. DeLucia, using Bloom's terms, supports "formative", rather than "summative", evaluation. "The use of formative evaluations after each separable unit or task in the learning process can do much to motivate the student to the necessary effort at the appropriate time."

Rosenthal and his associates studied the relation of teacher expectation to student performance. Dr. DeLucia wonders if the self-fulfilling prophecy, supported by their findings, might be arrested by witholding student records from teachers. Should we burden the teacher with the knowledge that a certain child has a history of failure? Should we burden the child with such a label?

Discussion

Moderator: Quentin D. Sprague
Director of Guidance
Cape Elizabeth (Maine)
High School

Reactors: Dr. Janice Cowsill, R.S.M.
Reading Consultant/Diagnostician
Title I ESEA - Newport, Rhode Island

Vincent J. Worden
Assistant Professor, State College at Bridgewater, Massachusetts

"Doctors have it good; they bury their mistakes. Teachers have it even better; their mistakes bury themselves." (Dr. Vincent J. Worden)

There was basic agreement with Dr. DeLucia's presentation. Sister Cowsill felt that local evaluation is essential, considering the varieties of community variables (language, home, etc.) She wondered if the state departments and/or agencies might offer their expertise in the development of local evaluative instruments and good research design. She cautioned the public about the assumpton of validity given to an "objective" test, the assumption that holds "if it's objective, it must be right."

The Sister and Dr. DeLucia discussed the Rosenthal findings. Sister Cowsill felt that if they are valid (and she does not yet grant their validity), we should educate teachers to be more professional, show them research indicating their biases, rather than throw out twenty years of development in the measurement field.

Dr. Worden was concerned with the practical problems of compensatory programs. Where do we find minority group personnel to staff the programs? Anne Anastasi (in an article published in 1953) indicated that children perform better on tests if they are administered by members of their own race. What are the implications of that for evaluative testing and for teaching? How can we adjust to the knowledge that the disadvantaged have little achievement motivation? Dr. Worden made a plea for early intervention. Present test content appropriateness for diagnosis is an issue that needs attention. He argued that providing for early intervention is more important than testing.

Must tests be administered <u>exactly</u> according to the Manual's directions? An audience member felt that the relaxed interaction between the teacher and child was severed by the requirements of publishers. The stilted, rigid instruction-giving process seems inappropriate to serving the disadvantaged.

Finally, all acknowledged the need for pure research on compensatory programs. Such research would require a control group, and few are willing to deny the benefits of compensatory practices to needy children to satisfy the needs of research. Therefore, basic research probably will not be funded by Title I.

COMPUTERIZATION IN RELATION TO TESTING AND EVALUATION

By DR. JAMES R. BAKER
Director of the Research and Development Center
Massachusetts State Department of Education

The three major functions or applications of computer technology in relation to testing and evaluation are: (1) administrative; (2) training and; (3) instructional. Computers can supply accurate analysis of student data and provide information for assistant superintendents and teachers, facilitating their efforts in the decision making process. The two primary types of data which could be derived are: (1) descriptive, e.g., pupil environment, nature of behavior, processes and; (2) data regarding the field of measurement, achievement and ability testing.

Valuable judgements cannot be made about a student unless a reasonably objective observation ensues from measurement information. Education has not arrived at a particularly sophisticated level of operation. Specifically, limited data are transmitted from teacher to teacher, school to school, school to college. Merely looking at a student's cumulative record shows a teacher very little about the student and his abilities. A listing of A's, B's, and F's is simply confusing.

In addition to giving evidence about a student's performance in a meaningful form, methods must be developed to provide viable avenues of interpretation of the effectiveness of teaching methodologies. The Federal Government is making great demands of Title I programs now and the present accumulated data base is inadequate to meet these demands. Computer technology could provide the required storage of descriptive materials and provide the information for effective analysis of a given methodology.

Computer applications operate at two levels. The first, assuming that there won't be any extensive trend toward computer assisted instruction, is concerned with the analysis of descriptive data about a student, recording measureable, transmittable data, detailing his educational progress. Further, such applications would provide an analysis of educational quality, a community's needs and advantages and the reasons for the expenditure for each student as compared with other school systems. A computer system would give a reliable inter-community cost comparison, e.g., average cost of test books, expenditure per pupil for guidance facilities, etc.

The second level is that of instruction. An example of the instructional component would be the C.A.M. Project (Comprehensive Achievement Monitering). In this program, students follow a curriculum set and have periodic examinations which are graded by computers, giving them a complete diagnostic report. The report includes comments about their strengths and weaknesses as well as suggestions for further studying or review. The C.A.M. Project, however, is only a minimal application of computer technology. With the appropriate staff, linguistic and instructional experts and computer programmers, programs for learning in accordance with each student's individual needs could be developed. With a combination of teacher and computer, the student could learn more effectively and more accurate, transferable assessments of a student, his abilities and weaknesses, could then be made.

Discussion

Moderator: Louis Gelsomino
Coordinator, Personnel and Pupil Services
Portsmouth (R.I.) School Department

Reactors: Dr. Robert D. Cloward
Professor of Psychology
Rhode Island College

Dr. John F. Espey
Editorial Liaison Officer
Test Department
Harcourt, Brace & World, Inc.

One of the primary concerns facing educational administrators is the per pupil cost of education. For school systems to initiate a greater use of computers, three primary problems must be dealt with. The Public must be convinced that the investment is legitimate, that it will not only facilitate but enhance the learning process. Each year, as it is now, superintendents have difficulty with budget approval. If the public can be convinced that computers have educational merit, acceptance will probably ensue. Closely related to this problem is a second consideration, suspicion and fear. Neither the role of the computer nor the role of the teacher, nor their relationship has been clearly defined. Does computerized education mean the replacement of the classroom teacher or will the computer function more as a resource agent to the teacher? The third problem is that hardware tie-ins and computers themselves, are not yet ready to be used on a large scale educationally, at least on an instructional level. There is no way at present to tie in Honeywell Computers with those made by IBM. For the most effective data bank, this must be accomplished.

The computer has been characterized as a very fast idiot. One must know how to use it because it only does what it is told to do. School systems which now have computers often don't use them to their full capacity, primarily because they do not know how to use them most effectively; they don't have the appropriate personnel. Because of this consideration and an examination of a computer's full potential, perhaps the role of "teacher" as it is presently conceived and the composition of a school's faculty will change. The programmer would become extremely important. However, he would not be effective alone. He would have to work closely with instructional and linguistic experts. The programs' primary purpose, naturally, will be to initiate, facilitate and continue the learning process. To do so, the language used must communicate the intent, goal, and purpose the instructors had in mind. Therefore, there will have to be extensive coordination in program development.

It has been suggested that with the advancement of computer technology in education, three catagories of teachers will evolve: pupil managers, subject experts and learning diagnosticians. The pupil managers would function in a supervisory capacity, instructing the students in the use of

the computers and assisting the children in whatever way necessary. The subject expert teacher would be one who enjoys learning. He would be responsible for task analysis and would work in conjunction with others in the development of programs. The learning diagnostician would function much as a guidance counselor. When a child demonstrates learning difficulties, he would deal directly with the student to determine whether or not the program meets the child's specific needs or if other intervening variables are interfering with the learning process.

The computer can give a more complete diagnostic report of a student's abilities and weaknesses. It can give a thorough diagnostic print out which would be interpreted by the student and teacher showing what needs to be learned and what was learned well. If information is to be transmitted from school to school, school to college, the computer provides the most efficient, comprehensive vehicle for this task. Further, as an instructional medium, its potential is just beginning to be considered.

TESTING AND ITS RELEVANCY TO THE SEVENTIES

By THOMAS BURNS
Acting Associate Commissioner of Education
United States Office of Education

There are many prophets of educational doom in the United States and much of their pessimism is warranted. However, they are quick to overlook what education has done well and seem to ignore the strong directionality that education is assuming toward more relevant programs and methods of learning. They also seem unaware at times of the large number of capable, creative teachers and administrators who are beginning to improve the educational process dramatically.

Part of the explanation of why education is not as valid or relevant as it should be is to be found in testing. We have misused tests and directed them to the wrong object: the student. If educational accountability has any meaning it is that the professional must be responsible for the products of his actions. Hence, the student isn't the failure, it is the school. As it is seen now, however, no teacher, administrator or counselor is responsible if a child fails. This is ludicrous and must end.

The growth of testing, evaluation and counseling during the past fifteen years has been phenomenal. Testing has increased at a rate greater than the school population. A student during twelve years will typically take five standardized achievement test series in three to nine subject areas. In 1954-55, 169,598 candidates took the College Entrance Board Examination. In 1968-69, 1,956,422 candidates took it.

The counseling profession has also experienced a similar rapid growth. In 1962-63, the secondary counselor ratio was 1:530. In 1967-68 it was 1:420, a growth of 26.4%. Government spending for counseling has also increased as a result of a national concern over the public schools. It had its origins in the late 50's: the cold war, the stalemate in Korea, and the orbiting of Sputnik. Suddenly the Russians were ahead in our race for survival. Our economic superiority had vanished and now our technological genius was in question. The intervention of the Federal Government into education had one single purpose: to identify, cultivate and promote talented students who could contribute to the growth of the United States. Testing seemed to be the best vehicle to meet this end.

The testing boom was assisted by the growing confidence in the social sciences and the alleged benefits that would come from the application of a scientific method to sociological problems. As the Government increased its demand for product assessment, educators increased their use of testing. Schools became extremely efficient at deciding which student was bright and worthy of college and which was not. They became channeling, sorting and certifying institutions, every bit as powerful as the draft itself. The past decade was an era which indicated that we must not penalize the bright student by putting him into a class of dullards. (Besides, the dullards would rather learn with each other, wouldn't they?) During these years of

the quantitative and qualitative change, we never once asked which school, which superintendent, which teacher, or counselor was the worst. Not once did we ask what kinds of teaching methods were the best.

We directed our questions about students to tests that couldn't provide the answers. We expected group tests that were unreliable with questionable validity to give us answers about specific individuals. The counseling profession has rigidified and strengthened this situation, rather than attacking its injustices. The remedy for this system is the application of the principle of accountability of the profession in the following ways. Counselors should thoroughly review and critique the tests being used in their systems by raising such questions as: do the tests have content validity? From what sources do they derive their concurrent validity? How reliable are they? Tracking should also be reviewed and critiqued, with a specific look at the consequences of tracking. The school system's vocational program should be studied to be certain that it is a real program, not just a dumping ground for those who are not going on to college. The opportunity for a student to move up in a tracking situation should be analyzed.

These are statements that are currently being considered in the Federal Office of Education. Perhaps these questions assign blame incorrectly, but the fact is that these are the kindest comments being made in Washington today. Already many have given up hope. For those who have not, this is the eleventh hour for education and education must stand up and be accountable.

FEDERALLY FUNDED PROGRAMS

By THOMAS BURNS
Acting Associate Commissioner of Education
United States Office of Education

Mr. Burns made his presentation very informally, preferring more group discussion and interaction.

He presented a summary of the legislation for the 1970 Appropriations with respect to the Elementary and Secondary Education of 1965. There are 27 federal programs at a cost of $2 billion, which is 50% of the allocation in the United States Office of Education. The funds going to education on the federal level amount to $10 billion, so in actuality, there is less than 50% control by the Office of Education.

The following is a summary of appropriations presented by Mr. Burns:

Title I.... Program for Disadvantaged-Elementary and Secondary
1970 $1.339 billion
1971 House Recommendation - $1.5 billion

Title II... Library Resources
1970 $42.5 million
1971 House recommendation - $80 million, with a strong emphasis on the right to read - every youngster will read before he leaves school.

Title III.. Innovative Programs
1970 $116,339,000
1971 $137,339,000 ($21 million additional)
$4 million - right to read
$17 million - guidance, counseling, testing

Title V.... Strengthening State Departments and/or Agencies of Education
1970 $29,750,000
1971 remains same

Title VI... Education for the Handicapped from $84 million to $105 million.
Title VII.. Non-English Speaking Education Program
1970 $21,500,000, a marked increase from $7.5 million in 1969.
Title VIII. Drop-out Prevention Program from $5 million to $8 million.
Appeal to Congress to increase to $15 million.

In addition:

Recommendations that the Pinpoint Disaster Program not be funded federally. The Commissioner has approved this recommendation. Cuban Refugees Fund - If any state has a 20% increase of Cuban refugees within three years, that state would be entitled to federal funds. Tidings Amendment - allows the carrying over of funds from one year to another.

Following this summary, Mr. Burns spoke of the delay in the 1970 appropriations. This delay which produced a tight time squeeze caused many problems associated with late funding, magnified also by unavoidable State Department and/or Agencies delays. Mr. Burns commented that due to the results of delays in fiscal year 1970, we have noticed a definite change. He thus concluded that 1971 allocations would be known prior to the start of the 1970-1971 academic year.

Discussion

Moderator: Richard P. Tardy
Director of Elementary
School Guidance
Concord, New Hampshire

Reactors: Dr. Catherine M. Casserly
Assistant Superintendent
Providence (R.I.) School
Department

Dr. Wallace Roby
Consultant, Compensatory
Educational Program
Connecticut State Department
of Education

In reacting to Mr. Burns' presentation, Dr. Roby commented that once equipment has been purchased, he sees no real effort to provide "data base." Mr. Burns agreed that the secondary and elementary levels have not yet learned to get greater milage from the computer. They only seem to use it for scheduling to date.

Dr. Roby then raised the question regarding the Tidings Amendment, "If funds can be carried over another year, will there be a cut in the following years' allocation?" Mr. Burns thereupon stated, "There is no cut in funds. The funds would be at least as much as the previous years' funds, or perhaps more. The idea is to completely spend the entire years' funds first, before touching the following year's funds."

Dr. Catherine Casserly recommended more funds for increasing school personnel. The major reason of failure, she believes, is that students have not been taught. They have merely been exposed to ideas, but not to teaching. In discussing accountability (i.e. doing what you said you would do relative to defining and carrying out objectives), Dr. Casserly pointed out that accountability goes back to behavioral objectives; i.e. "do your thing" that will really teach children. There is a great shortage of evaluation personnel, and she sees the need for evaluation in education by means of follow-up and follow-through.

In response to the question as to whether there is any possibility for adult education beyond the eighth grade receiving federal funding, Dr. Casserly replied that you cannot use Federal monies for adult high school diplomas.

Presently, the Office for Economic Opportunity pays for high school equivalency. Mr. Burns pointed out that such a possibility might take a new piece of legislation. Should adult education beyond eighth grade be a federal funded responsibility? Or should not this be considered a good cause for local tax payers?

Mr. Burns concluded the discussion by offering the following three points:

1. The need to concentrate on good planning rather than working within the budget. (He believes in the systematic approach-building for the sake of good planning, rather than getting locked into a budget.)

2. The right of every youngster to obtain the education he wants and needs, be it vocational or academic - e.g., at age 16, a boy should be able to go out and run his own gas station, if this is his wish.

3. The necessity for the extension of the domain of Pupil Personnel Services to include guidance and counseling relating to helping youngsters while at the same time providing a service for them.

DISCLOSURE OF TEST RESULTS

By THOMAS P. NALLY
Professor of Education, University of Rhode Island

Testing is at best a complex business fraught with dangers of misconceptions, misinterpretations and misuses. The goals are noble, the means still open to many questions of a psychological, sociological, and technological nature. With this in mind, any consideration of the disclosure of test results must consider at least four areas: (1) the invasion of the individual's privacy; (2) the protection of school personnel from possible legal action; (3) the parent's right to information contained in pupil records and (4) the public's right to know on standardized measures the results of the performance in the schools it supports.

The debate over privacy seems to center around two issues, the consent of the student and the confidentiality of the information. Test cases have not as yet reached the courts although the Association of the National Council on Measurement in Education is attempting to get a half dozen subpoena-of-data cases on the docket. The American Civil Liberties Union supports a policy which is designed to protect the privacy of students with regard to prospective employers. They state that the school personnel may assess the student's ability to perform but not questions of loyalty, patriotism or morals. The American Psychological Association in its statement on psychological assessment and public policy states emphatically that "an individual should be protected against unwarranted inferences by persons not equipped with the requisite background of knowledge."

The second issue, that of protection of school personnel from legal action, is quite clear so long as such personnel do not maliciously slander the pupil. Truth is the best defense for defamation and there is further legal protection under the concept of privileged information. Indiana actually has a statute protecting school personnel.

The third issue, parents right to know, is also quite clear-cut legally, even though the argument over disclosure continues. It is contended that parents can set enlightened goals for their children if they have the appropriate information. In contrast, school personnel opposing the practice of telling parents each child's test scores have claimed that many parents do not understand how to interpret test results wisely and frequently distort the meaning of scores.

The fourth area of discussion is by far the most controversial. The issue of disclosure of test results on a system, state, region, or nation-wide basis has generated far more heat than light. The early attitudes of school administrators are revealed in the American Association of School Administrators position of opposition on National Assessment. True concerns certainly can be generated for the problems which well might arise from the misinterpretations,

misconceptions and misuses made of the results of testing programs by those who are partially uninformed or misinformed as well as those with ulterior motives. The arguments for releasing test results are: (1) to increase the accountability of the public schools; (2) to identify inadequate performance attributable to inadequate programs, etc.; (3) administrative leadership might be created as educational leaders seek to educate the public regarding proper interpretations of a system's scores; (4) rapidity of change might well be encouraged; (5) characteristics of various populations might well be highlighted.

The arguments against disclosure are: (1) the disclosure of test results could lead to a dictation of curriculum by test makers; (2) in many instances, the tests do not take into account the localized character of educational programs; (3) disclosure will create excessive pressure on pupils, teachers and administrators to achieve as measured by tests; (4) disclosure may lead to unwarrented and unrealistic comparisons of schools and/or school systems; (5) the limitations of tests and testing programs are well known to professionals in the field of measurement. Interpretations are complex. Disclosure must be accompanied by an extensive educational campaign if misconceptions, misinterpretations and misuses are to be avoided.

Discussion

Moderator: Shirley M. Reid
Counselor,
Rutland (Vermont)
Senior High School

Reactors: Mrs. Ann Hill
Director of John Hope
Settlement House
Providence, Rhode Island

Dr. Ellis D. Tooker
Superintendent
Hartford (Connecticut)
Board of Education

Learning is colored by emotions on the part of the student and the teacher. Children aren't usually judged accurately by tests and teachers are often psychologically tricked into certain methods of teaching. If a teacher has been told that a child has tested low, he will teach with little expectation. This is unfair to the student. Testing, as it is presently conceived, is geared to the middle class student and has a definite tendency to freeze individuals into specific achievement levels. The disadvantaged student suffers the most from the instant, seemingly immutable categorization. Since he does not perform well on tests couched in middle class terms, he is labeled as dull, unable and inferior. In many instances, when test scores are ignored, such a student can begin to excel and develop intellectually since he is not constantly reminded by another achievement test score that he simply does not have the abilities to perform well.

Tests have serious limitations which should be examined closely. Test scores are generally viewed by the public and by many teachers as final, definite, discrete and unchangeable. It is not the test maker who has made this assumption. It is the individual who works from little real knowledge of tests who generates the misuse of tests and causes score disclosure to be a risky venture.

Private industry as well as the general public has come to believe that there is something magic about testing. Industry sees it as a possible answer to all of their "employment ills." However, standardized tests were never intended to answer all questions about an individual. Global intelligence cannot be measured. As testing is now constructed certain kinds of abilities can be examined that are necessary for particular aspects of education and success. Creativity, some attitudes, physical abilities, motivation, talents in art and music, however, pose a problem for the tester since it is extremely difficult to measure such aspects accurately.

A problem associated with the disclosure of test results is the lack of interpretational abilities by the teacher and the parent. If a teacher assumes that a test score is 100% accurate and teaches in accordance with the score, he is making a mistake. If test results are to be disclosed, the scores must be given out with an accurate interpretation of their meaning. Not all teachers have the necessary knowledge to understand a score on a standardized test and it is unlikely that most parents do. Tests as presently constructed, fail to assess the child as a whole; his mind, his body, his emotions and those factors which play upon him.

NORMS: FACT OR FANCY

By DR. WALTER N. DUROST
Educational Consultant, Test Service and
Advisement Center, University of New Hampshire

Standardized tests, in the modern sense of the word, are relatively young. The procedure of obtaining norms has developed enormously from a rather informal, unscientific approach to a highly complex process of stratified random sampling of large numbers of pupils.

The use of mental ability measure along with the achievement test provides assurance that the norm samples tested are representative of the national population on measured mental ability.

New types of norms have been developed which are in many ways more serviceable than those previously available, especially stanines. Great encouragement also has been given to the practice of establishing local norms using stanines.

Grade equivalents, widely used by popular demand, are largely discredited by the professionals in Measurement. Their main value is in the interpretation of averages, not pupil scores. The assumption that a 12-month gain really occurs in the 9 or 10 month school year, the unwarrented extrapolation to levels not covered by a given battery, the lack of equality of units from level to level and test to test, all argue against its use. Norms making possible comparisons with peer groups are much superior.

Criterion reference testing (or test interpretation) is being talked about as if it were new. The adage, "Anything worth doing, is worth doing well" (perfectly?) defines the philosophy of its proponents. When an hierarchy can be shown, mastery of a skill prerequisite to success at a higher level makes sense; 100% mastery of much of the trivia in the curriculum does not.

The advent of optical scanners and computers has made the processing of tests a thousand times more rapid but the logistical problems of transporting the materials to a given testing center and of distributing the results still cause trouble.

A renewed emphasis on item analysis in the achievement test area where it is relevant has thrown the teacher's attention back to the item as the essential building stone of any test.

Testing and related methods of objective evaluation have more relevance than ever in the area of the individualized curriculum. The emphasis of the day is on accountability and the need is for more precise and accurate ways of evaluating educational outcomes in order that the methods of teaching, the content of instruction, and the organization of the schools may become more effective.

Discussion

Moderator: Clarence H. Steinberger
State Guidance Consultant
Connecticut State Department of Education

Reactors: Dr. John A. Finger, Jr.
Director, Educational Services Center
Rhode Island College

Dr. Joseph A. Whelan
Curriculum Director
North Providence (R.I.) School Department

In the discussion it was pointed out that teachers need to get involved in the testing situation and not put test results aside. Test results are extremely important at the time they are received, not days, months, or years later.

Programmed Instruction has not proved to be differentiated to the class.

The process of education is very complicated. We need to use every means available to inform us of what we are or are not accomplishing. Testing can give us a great deal of objective data, but in the final analysis, the child must satisfy his instructors and the school that he has learned. It is the instructor's responsibility to communicate with the children so effectively that they will learn those things on which are placed the most value.

Dr. Durost believes the "normal curve" will never be repealed. Accountability has to be directed to teachers and curriculum makers, as well as to children. We must present the child with a relevant curriculum. There must be continuing instruction in basics, as a child has to know how to read and work with numbers to cope with today's world.

TESTS: WHO OR WHAT IS BEING EVALUATED

By DR. C. THOMAS SKAGGS
Assistant Professor, College of Education,
University of Maine

The processes of measurement and evaluation are a foundation on which stand the functioning of man and his society. The processes have been in use since the earliest introduction of man onto this planet. The basic principles of measurement and evaluation, although not formally organized into a science or a studiable field of knowledge, were employed by even the earliest civilizations. As individuals and their society became more complex there were increasing demands placed on their "measurement techniques."

When we are confronted with any decision, we apply some type of "measuring instrument" to the situation and obtain an estimate or impression of the status of the circumstance. On the basis of our past experience and our assessment of the situation, we evaluate the total happening and reach a decision. Some thought on the part of each of us will produce many examples of our routine use of measurement and evaluation in the decision-making processes we employ in our own behavior. This is an informal, non-scientific, often unconscious application of measurement and evaluation techniques. We test the circumstance and make an evaluation of the test results in order to aid us in determining a course of action.

A second phase of the measurement and evaluation scene is our use of the techniques in an organized or structured manner; the development and synthesizing of our intuitiveness into a defined science--the science of psychometrics. As defined by Lee J. Cronbach, this term refers to the obtaining of a numerical estimate of some aspect of performance at a given time. A most familiar application of this science is by those of us who share responsibility for the American educational system.

In general, the most common application, which is achievement testing, has two branches: (1) teacher-made classroom tests and, (2) professionally developed standardized tests. While both types of tests are designed to assess basically the same variable, achievement, the two are not interchangeable in a school testing program. Each has its own use and misuse. It is generally agreed upon that the primary purpose of testing and obtaining test results is to aid in the improvement of the teaching-learning process. Informed use of test results in this manner is certainly a worthy expenditure of effort and a worthwhile goal. Unfortunately, however, test results are often misused. In order to look good, teachers sometimes desert the approved course of study and teach for the test. Administrators use test results to evaluate teachers and teacher effectiveness. Tests tend to freeze the curriculum in the patterns approved of by the test makers. The results of national and state programs are unfairly used to compare school systems. Testing programs promote undesirable competition among schools.

Students express the following points of view in relation to testing. Tests are not used properly in the classroom, nor do they directly help the student. The instructor does not use the test results to improve his course. Most tests are very poor. They don't measure anything beyond verbal level of learning. Finally, teachers feel that they must give tests to help them in assigning course marks rather than using more appropriate criteria.

In conclusion, when asked "Tests--Who or What is Being Evaluated?", in the broadest and most responsible sense of the word, the implications of the science of measurement and evaluation are so far reaching that nearly every phase of our total being is under evaluation. The practices of testing in today's society is an aspect which we all must responsibly evaluate and be certain of where it is leading us in our very competitive American culture.

Discussion

Moderator: Sister Mary Edward, R.S.M.
Assistant Superintendent
Diocesan Schools, Rhode Island

Dr. Dorothy D. Moore
Instructor in Education
University of Maine

Dr. Robert W. Read
Associate Professor of
Counselor Education
Northeastern University

General audience discussion centered around teacher abuses of tests. Very often, teachers who have never taken a test and measurement course in college try to measure the achievement of their students. They are the most guilty of misusing tests. Parents in conversation with teachers, frequently request information about their child's test scores so that they can use them merely as status symbols thrown out casually in conversation at cocktail parties. Even though no test measurement is absolute, scores are usually treated that way.

Teachers are, on occasion, guilty of teaching the test but it is their fault only indirectly. Parents and school administrators apply much pressure to have their children excel on standardized tests in relation to national norms. If the teacher does not teach the test and the children do not do well, the implications for the teacher are clear. In response to this, Dr. Moore agreed that tests aren't usually used to help the student. More should be done with the tests than scoring and filing them. They can and should be used as diagnostic tools which would indicate to both the teacher and the student a study directionality. Diagnostic test results point up areas of strengths and weaknesses and indicate clearly what needs to be done or redone and what was done adequately.

The audience and panelists concurred that curriculum does often end up being directed by standardized tests. Such a dictation creates a middle class bias in curricula since most standardized tests are geared for the middle class student. In turn, when a child from a lower socio-economic level is given a standardized test, he invariably comes out at the bottom since the test is not appropriate for him. However, such achievement tests present fewer problems for students than teacher-made tests. Teachers aren't adequately prepared to develop and interpret the scores. Further, the self-made tests are plagued with low reliability and validity coefficients.

There is no point, as Dr. Read emphasized, in testing unless it has implications for the future, unless the teacher does something with the information to help the student. Tests will always have limited value, however, since their predictive powers are generally low. Tests are used to evaluate teaching, the individual school and the school system. A low score on a test might indicate that the student has failed in his task, but it might also be an indication that teacher, or perhaps the curriculum, are at fault, not the student.

THE JENSEN REPORT

By DR. PAUL B. CAMPBELL
Office of Research and Statistics,
Pennsylvania State Department of Public Instruction

It is quite common for an individual to hear what he wants to hear and disregard everything else. This is the fault with the majority of people who concern themselves with Jensen's findings. They take excerpts out of context, and take word-of-mouth versions, or make-over generalizations, which are not helpful to the solution of the problems of providing ample opportunity for self-fulfillment to all mankind.

Is there such a thing as individual difference in ability? If so, can it be utilized? Should it be changed? Are there ways of educating which really take into account existing differences?

The investigation of the above questions has challenged the beliefs and hopes of many; however, because of this challenge, there is no reason to cease the investigation, or to personally attack Mr. Jensen, but rather the investigation must continue to bring about more hypotheses and further examinations.

Exactly what does Jensen say?

First, he says that compensatory education has been tried and has failed; and to provide evidence in support of the above statement, he cites the Civil Rights programs saying that none appear to have raised significantly the achievement of disadvantaged students as a group. Jensen's purpose in bringing up this matter of compensatory education was to provide evidence to show that heritability accounts for from 70% to 90% of the variance of the obtained scores in specific populations.

The critics, in general, agree that there is high likelihood for a genetic base for intelligence. The differences appear to focus around the quantifiability of the relationship, the possibilities of modification by environmental manipulation, and the role of genetic factors in explaining the differences in achievement and I.Q. between populations which differ on socioeconomic, ethnic, or racial descriptions.

Jensen's critics do not propose alternative values for the heritability of intelligence but remain agnostic about the causes for differences.

Jensen reports numerous studies of manipulating the environment, which achieved temporary effects. While somewhat impressive, the final scores reported were still within or near the standard error of estimate based upon average heritability in most cases. What Jensen is saying is that environmental manipulation can account for only a small portion of I.Q. variance and that part of the accountable portion is outside the province of society's conception of educational practice (e.g. nutrition, number of children per family, amount of individual attention).

Jensen cites evidence from several sources to define the relationship between I.Q. and socioeconomic status. Jensen sees this relationship as associative, and not causative. Socioeconomic status is merely a proxy measure of I.Q., he believes.

Intelligence is defined by Jensen as only one aspect of what he calls mental ability. He then reports data to show that intelligence so measured is substantially lower among blacks than among whites. Needless to say, he meets attack from his critics, but they offer no really helpful comments and leave the issue unresolved.

The majority agree that there is some genetic basis for intelligence, and we must face the issue of individual differences with the knowledge that everyone can not be expected to perform at the same level. Each individual must be given the opportunity to develop his potential to its fulfillment. The goal of education becomes diversity of opportunity, with maximum assistance to each individual to find for himself his optimum fulfillment. Jensen believes that there is potentially much more that can be done to improve school performance through environmental means than through the changing of I.Q. per se. The focus should be upon achievement rather than changes in intelligence.

The issue becomes one of allowing Jensen to continue his work in helping mankind understand better the role of genetics and the kind of learning ability that may be genetically determined.

Discussion

Moderator: Dr. Charles R. Dolan
Associate Professor
Counselor Education
University of Bridgeport,
Connecticut

Reactors: Dr. John A. Finger, Jr.
Director, Educational
Services Center
Rhode Island College

Jarvis D. Jones
Assistant Professor of
Secondary Education
Rhode Island College

Dr. Robert W. Stoughton
President, AMEG
Chief, Bureau of Pupil
Services
Connecticut State Depart-
ment of Education

There is enough intelligence in nearly all of humanity to handle most of the tasks to which we have to address ourselves. For example, everyone should be taught to handle the number system, to read, etc. Those who do not achieve this level have what he calls a physiological impairment.

There is a wide range of abilities and talents in many different areas, and well over 50% of the population is at the top of the scale in some factor of intelligence. We have individual differences but they are not necessarily limiting to people. We can expect people to accomplish successfully general operations within society and very well in a specific area which is suitable to the individual.

Dr. John A. Finger: We cannot settle the issue of intelligence and the heritability of intelligence. We cannot say how much gain we can expect from the best compensatory education program, because these programs have not been effective, and we should have been able to make this prediction. Compensatory education programs must be improved and we have to try to determine what results we should obtain from good compensatory programs. It is a question of discovering where the effective ones have been, and then studying them.

We tended to classify by race rather than socioeconomic differences and we need to examine our whole way of looking at this problem. We assume that race is the only way to classify, rather than realizing that socioeconomic reasoning has the solution. Environment can have such a major influence on some children while not on others.

Dr. Robert Stoughton: Heredity may be the more important factor in intelligence which is not the same as mental ability. It may be more important than environmentalists assume and there may be racial differences in mental characteristics. Through factors other than intelligence which affect achievement and success, we can do much to improve performance through environmental means.

The important point Jensen was emphasizing is to treat each person as an individual; that variables such as social class, race, national origin, are correlated so imperfectly with any of the valid criteria on which education or employment decisions should be made that they are irrelevant as bases for dealing with individuals.

Socioeconomic status and environmental forces can depress or elevate an individual's ability to advance. Jensen doubts that educational intervention will significantly affect intelligence, as he defines intelligence; however, educational intervention can affect achievement. Intelligence is an irrelevant hurdle; there are other evidences of capacities to perform.

The term I.Q. is emotionally charged. Jensen's recognition of other than mental factors in success gets lost by laymen, who insist on believing the I.Q. is unchangeable. High heritability does not mean that environment cannot be important - it means it has not been important.

Using tests as prediction are merely excuses for human failure. Tests help us determine the intervention we should use.

Question from floor - What kinds of interventions will succeed?

Comments from floor - Personalities of people involved in education are more important than structured programs. Our perceptions of youngsters in essence determine their behavior. It is important to be concerned with the cognitive and affective elements of persons. People are the most important elements of learning and growth, rather than structured programs.

STATEMENT OF PURPOSE

NEW ENGLAND ASSOCIATION FOR MEASUREMENT AND EVALUATION IN GUIDANCE

New England AMEG is an affiliate of the Association of Measurement and Evaluation in Guidance, a division of the American Personnel and Guidance Association. The primary purpose of this organization is to provide leadership in educational measurement and evaluation in New England. Further purposes indicated in the By-Laws are to:

a. provide a forum for the discussion of problems related to measurement and evaluation within the guidance area.

b. identify problems and concerns relating to the use of tests and other evaluative devices in the personnel and guidance area.

c. stimulate research relating to tests and their use in guidance.

d. promote the effective training of personnel in the use and interpretation of measurement and evaluation devices in guidance and personnel work.

NEW ENGLAND A M E G / Officers, 1969-70

President: Prof. Francis E. Dunn, Director of Educational Measurement, Brown University

President-Elect: Mr. Richard P. Tardy, Director of Elementary School Guidance, Concord Union School District, Concord, N.H.

Secretary-Treasurer: Prof. Vincent J. Worden, Bridgewater, Mass., State College

Membership: Dr. Charles R. Dolan, Associate Professor, Counselor Education, Colege of Education, University of Bridgeport, Bridgeport, Conn. 06602

Publications Chairman: Dr. Robert W. Read, Associate Professor, Counselor Education, College of Education, Northeastern University, Boston, Mass. 02115